基于用户决策机理的个性化推荐*

林鑫[1,2] 桑运鑫[3] 龙存钰[2]

[1] 中国科学技术信息研究所 北京 100038 [2]华中师范大学信息管理学院 武汉 430079

[3] 武汉大学信息管理学院 武汉 430072

摘要：［目的/意义］针对基于内容的个性化推荐策略，提出资源特征选择与权值计算优化策略，从而改善个性化推荐的效果。［方法/过程］构建基于用户决策机理的个性化推荐模型，模型以用户决策机理为背景知识进行资源特征的选择、用户兴趣模型的构建与语义表示、用户决策函数构建。为验证模型效果，以4 748位用户的观影数据为例进行实验，实验以向量空间模型为参照模型，P@N为评价指标。［结果/结论］实验结果显示，在N取值为5、10、20、50、100、200的情况下，基于用户决策机理的个性化推荐模型效果都显著优于向量空间模型，从而验证模型的有效性。

关键词：决策机理 基于内容推荐 个性化推荐

分类号：G203

DOI：10.13266/j.issn.0252-3116.2019.02.012

近年来，作为一种解决信息过载的有效手段，个性化推荐受到了广泛关注[1-3]。其中，基于内容的推荐（Content-Based Recommendation）是一种较为常见的实现思路，其取得良好效果的关键是合理的资源特征选择与权值计算[4]。从基于内容的个性化推荐基本原理出发，理想的资源特征应该既是用户决策时的参考因素，又能够实现用户感兴趣资源与其他资源的区分；理想的权值计算策略则既需要反映单个资源特征对用户决策影响的大小，又需要在多特征融合时契合用户决策机制。但以往的研究过于关注所选资源特征及其权值计算方法能否实现用户历史感兴趣资源与其他资源的区分，而对其是否契合用户决策机理关注不足。一方面可能导致所选资源特征未必是用户决策时关注的因素；另一方面也可能导致所选资源特征在权值计算中未得到合理处理，进而影响个性化推荐的效果。为解决这些问题，可以将用户决策机理作为个性化推荐策略设计的背景知识，以此指导资源特征的选取、用户兴趣模型的构建及特征权值的计算与融合。

1 相关研究

近年来，国内外学者围绕电子商务与信息消费中的用户决策机理进行了多方面研究，并已将用户决策机理初步应用到基于内容的个性化推荐中，下面对相关成果进行综述。

1.1 电子商务与信息消费中的用户决策机理

决策机理研究在心理学、经济学等相关领域已经有多年历史，形成了包括理性决策模型、有限理性决策模型、前景理论、偏好构造理论等在内的多个决策理论模型[5]。近年来，随着互联网和电子商务的发展，国内外学者围绕网络购物和信息消费情境下的用户决策机理进行了探索，并取得一系列成果。李宗伟等从整体出发对影响消费者在线购买决策的因素进行了研究，认为主要因素包括商品价格、商品销量、卖家信用等级、卖家服务评级、卖家开店时间、在线评论长度[6]。围绕智能手机购买决策中的影响因素，K. L. Lay-Yee等[7]和J. Sujata等[8]分别进行了研究，前者认为主要因素包括产品特色、便捷性、品牌、价格、熟人影响等，后者则在因素归类上略有区别，将其归纳为技术、硬件、基础因素、品牌、价格5类。在数字图书馆信息资源利用决策中。S. Joo等认为用户的有用性和易用性感知，以及信息资源质量（可获得性、可信度、范围、新颖性和格式）显著影响用户的最终决策[9]。查先进等

* 本文系国家社会科学基金青年项目“社会网络中基于用户认知结构的知识标注研究”（项目编号：17CTQ024）研究成果之一。

作者简介：林鑫（ORCID：0000-0003-0318-8160），讲师，博士；桑运鑫（ORCID：0000-0002-0026-4319），本科生；龙存钰（ORCID：0000-0001-5096-7461），本科生，通讯作者，E-mail：345705868@qq.com。

收稿日期：2018-02-11 修回日期：2018-06-11 本文起止页码：99-106 本文责任编辑：徐健

则将影响因素分成直接影响因素和间接影响因素，前者包括信息有用性、对数字图书馆的依恋；后者包括数字图书馆的信息质量、数字图书馆的信源可信度、数字图书馆的声誉三个方面，它们通过信息有用性间接产生影响[10]。吴江和周露莎采用回归分析的方式对影响用户购买网络健康信息服务决策的因素进行了研究，并将其归纳为医生职称、医院等级、评论数量、好评率、感谢信数量、诊后报到患者人数几个方面[11]。

1.2 用户决策机理在基于内容的推荐中的应用

除少量研究之外，如选择频率中心、短时平均能量、过零率、MFCC、带宽等作为音乐推荐的特征[12-13]，其他基于内容的推荐研究都或多或少地考虑了用户决策机理。概括起来，可以将其分为特征选择中考虑用户决策机理和权值计算中考虑用户决策机理两类。

（1）特征选择中考虑用户决策机理。这类研究在个性化推荐模型构建中选择对用户决策有影响的因素作为建模特征，但在权值计算环节却将各类因素混在一起，不加区分，较为典型的就是向量空间模型。例如，面向大学生的图书推荐中将用户专业、年级、图书主题、作者等作为建模特征[14]，电影推荐中将类型、导演、演员等作为建模特征[15]，音乐推荐中将语言、民族、文化、位置、风格流派、歌手等因素作为建模特征[16]。

（2）权值计算中考虑用户决策机理。该思路是指在个性化推荐中，首先计算各个特征的权值，然后根据用户决策中这些特征的作用方式对其融合，生成最终推荐结果。例如，基于情境的个性化推荐中，首先分别考虑用户对资源本身的兴趣和情境因素，在此基础上对二者进行融合获得综合权值[17-18]；李江等在学术评审的专家推荐研究中，首先分别计算候选专家与评审对象的专长吻合度、学术影响力与社会关联值，在此基础上将3个维度权值的乘积作为衡量推荐与否的依据[19]；杨程等在进行面向开发者的开源项目推荐研究中，首先分别计算候选项目的流行度、与用户技术能力的相关度、社交关联度，然后对其加权求和，生成推荐列表[20]。

总体来看，国内外学者围绕不同情境下的用户决策机理进行了多方面研究，而且用户决策机理融入到个性化推荐中已经得到了一定程度的认可，并证实了其有效性。但以用户决策机理为主题的研究往往只注意到了其在市场营销方面的应用价值，极少提及其在个性化推荐中的应用意义。而个性化推荐中的用户决策机理应用还处于自发阶段，缺乏系统性：①特征选择往往基于研究者的经验、观察进行，而非从用户决策机理出发，自上而下的进行特征选择或抽取、挖掘，可能会导致所选特征集合无法全面涵盖用户决策的影响因素，如文献[19]提出的专家推荐策略中未考虑工作态度等因素[21]；②在特征间关联关系分析中，已有研究多是从自身经验或逻辑分析出发，而非以用户决策机理的系统研究为基础，由此就可能导致因素间的关系分析有偏差，尤其是特征较多时，进而影响特征融合的效果。针对以上两个问题，笔者首先构建一个基于用户决策机理的个性化推荐通用模型，立足于用户决策机理进行特征选择，用户兴趣模型构建与决策函数拟合策略设计，并在此基础上以电影推荐为例验证模型效果。

2 基于用户决策机理的个性化推荐模型

在基于用户决策机理的个性化推荐中，用户决策机理的定位是作为推荐策略设计的背景知识指导资源特征的选择、兴趣模型构建与语义表示、决策函数生成，使得构建的推荐模型可以更好地拟合用户的实际决策，从而改进个性化推荐的效果。模型构成要素及其关联关系如图1所示：

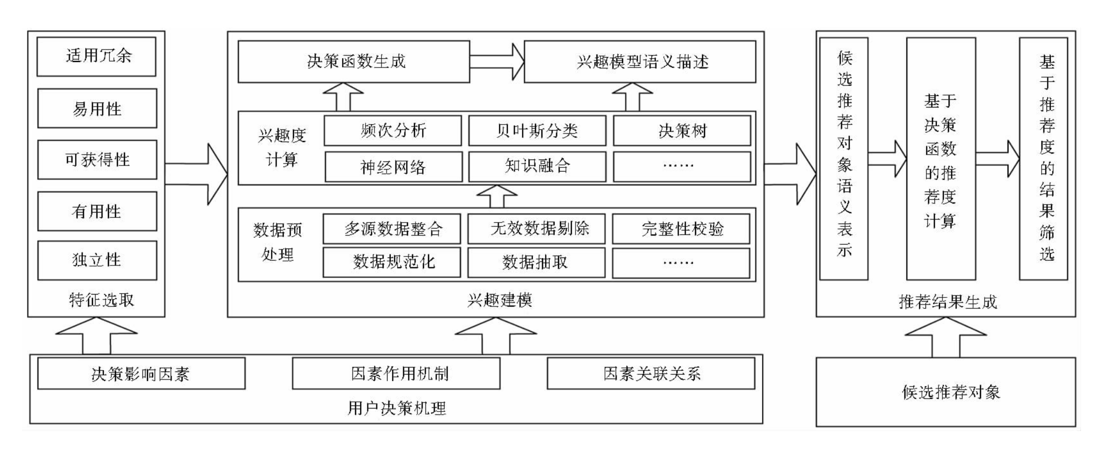


图1 基于用户决策机理的个性化推荐模型

2.1 用户决策机理分析

用户决策机理分析是后续几个模块运行的基础,也是影响推荐效果的关键因素。其核心任务是,针对拟推荐的产品或服务,厘清影响用户决策的主要因素、作用机制、决策过程中各因素间的关联关系等。在分析实现上,心理学领域已经提出了多种行之有效的方法,包括元分析法、实验法、观察法、问卷法、访谈法、回归分析法等[22],在具体分析中可以结合拟推荐产品或服务的实际情况进行方法的选择。

值得指出的是,在分析过程中需要特别关注以下几个方面:①需要区分直接影响因素和间接影响因素。②需要实现决策因素的具体化,至少需要保障全部间接影响因素以及不存在间接影响因素的直接因素的具体化。③在因素的作用机制分析中,一方面需要特别关注多取值影响因素的作用机制,以电影演员因素为例,需要特别关注一部影片有多个用户喜欢或讨厌的演员时,其作用是正向的还是负向的,影响程度如何变化;另一方面还需要关注因素间的关联关系,即用户决策时的效用函数形式,最为常见的是线性函数,但也可能存在非线性函数形式,如 U 型函数。④受个体决策的影响,可能存在多种决策模式,即不同用户决策中考虑的因素是有差异的,各因素间的交互关系也可能是多样化的,在分析中应涵盖常见的决策模式。

2.2 特征选取

资源特征是兴趣建模与推荐结果生成中的直接处理对象,因此为实现个性化推荐中对用户决策机理的拟合,需要在特征选择环节就将用户决策机理融入进来。①特征选择应采用自顶向下的思路,即从影响用户决策的因素出发,寻找能够反映该因素的特征或特征组合;②特征选择中应坚持独立性原则,即尽量避免因素间的耦合,但需要依据用户的决策机理判断因素间是否耦合,而非其客观上是否存在相关关系,如影片的导演、主要演员等主创作者信息与其国家地区强相关,但用户决策中一般将其视为两个因素,因此在特征选择中需要全部纳入进来;③对于同一个直接影响因素,用户可能采用不同的间接因素(组合)进行判断,同时单个特征也可能存在缺失值问题,因此在进行特征选择时可以保持适当的冗余。

此外,为保障特征选择的合理性,还需遵循以下几个原则:①有用性,即所选特征至少与一个影响因素有关联;②可获得性,即所选特征对应的数据是能够获取的,否则该特征无法发挥作用;③易用性,资源特征需以量化的形式进行利用,为降低实现难度,需要选择易于量化的特征。

2.3 兴趣建模

该模块的作用是立足于所选资源特征及用户决策机理,根据用户的历史行为及相关资源的信息将用户兴趣偏好以决策直接影响因素-兴趣度的形式进行显性化,并进行语义描述。实现上可以分为数据预处理、兴趣度计算、决策函数生成、兴趣模型语义描述 4 个环节。

(1)数据预处理。其任务是以用户历史行为、资源相关数据为基础,结合所选资源特征,将数据处理为完整的、规范的、便于深度加工的形式。在该环节中,需要进行的操作包括多源数据整合、无效数据剔除、数据完整性校验、数据规范化、连续数据离散化等[23]。此外,建模所需的部分特征可能无法直接获取,因此在数据预处理中可能还需要应用数据抽取、统计分析、分类聚类等数据挖掘方法。

(2)兴趣度计算。为便于后续利用,兴趣度计算中需要改变以特征为粒度的计算方法,以用户决策机理分析出来的直接影响因素作为基本单位。因此,在计算方法上,对于包含多个间接影响因素或由多个特征方可判断的直接影响因素,除了常用的频次分析、加权统计、贝叶斯分类、决策树、神经网络等方法外,还需要采用知识融合的方法加以处理。

(3)用户决策函数生成。在该环节中,首先需要根据用户决策机理中各影响因素间的关系确定函数的基本形式,进而采用回归分析法利用用户的历史行为数据进行决策函数的拟合。值得指出的是:一方面,为获得最佳用户决策函数,若存在多种可能的决策模式,需要对多种用户决策模式逐一进行拟合与比较筛选;另一方面,用户决策函数确定后,还需要视情况对用户决策机理中的影响要素进行调整,剔除无关要素,从而获得个性化决策模型。

(4)兴趣模型语义表示。鉴于各影响因素在用户决策中扮演着不同的角色,因此为便于推荐结果生成环节的利用,需要将用户兴趣模型进行语义表示。具体而言,以用户个性化决策模型中的所有直接影响因素作为语义表示的框架,从而将用户兴趣模型表示为一个高维向量空间,每个维度对应一个用户决策的直接影响因素,而每一个决策直接影响因素可以表示为由二元组(因素取值,兴趣度)组成的向量。

2.4 推荐结果生成

推荐结果生成的核心是基于用户决策机理,计算资源与用户兴趣模型的匹配度,并在此基础上挑选一

部分展示给用户,以避免带来新的信息过载。其实现包括候选推荐对象的语义表示、基于用户决策函数的推荐度计算、基于推荐度的结果筛选。

(1)候选推荐对象的语义表示。与用户兴趣模型的语义表示相似,候选推荐对象的语义描述也需要以用户决策中的直接影响因素作为框架要素。如果要素存在直接对应的资源特征,则根据其取值生成该要素的取值;如果不存在直接对应的资源特征,则需要根据原始资源特征与要素间的映射关系对其加以处理。

(2)基于用户决策函数的推荐度计算。在该环节中,首先需要利用向量空间模型分别计算资源与用户兴趣模型中各要素的相似度,在此基础上,根据用户决策函数进行各要素权值的融合,生成最终的推荐度。

(3)基于推荐度的结果筛选。在计算各候选推荐对象推荐度的基础上,可以根据具体应用需求筛选一部分作为最终推荐结果。较为常见的筛选方法是Top N法和阈值法,其中,Top N法是指将各资源按照综合权值进行排序,取排名最靠前的N个或N%个;阈值法是指设置一个综合权值阈值,将大于该阈值的资源都展示给用户。

3 实验

为验证模型的有效性,选择电影为对象进行实验,并以国内知名的电影网站豆瓣电影(https://movie.douban.com/)作为数据源。同时,为便于评价模型效果,选择常用的向量空间模型作为对照实验中的推荐模型。

3.1 样本数据

样本数据采用的是笔者于2014年11月20日–12月15日期间采集的一份包含830 682位豆瓣用户观影记录的数据集。采集内容包括:这些用户的全部观影记录,字段包括用户ID、影片URL、观影时间、添加的标签;样本中所涉及的101 486部影视作品的基本信息,包括URL、片名、导演、演员、编剧、类型、制片国家、上映时间、评分、观影人数、集数、片长等。

在获得基础数据后,首先剔除电视剧、综艺节目等非电影类数据、缺乏上映时间字段的数据,以及对应的用户观看记录。在此基础上,进行样本集构建:首先从2014年5月3日–14日期间观看至少1部影片的用户中,随机抽取5 000位作为初始样本;其次,为避免数据过于稀疏,剔除了2014年5月3日以前观影少于20部的用户,剩余4 748位用户。进而,将这些用户在2014年5月3日以前的观影记录作为兴趣建模的数据集合,共620 612条记录;2014年5月3日–14日的观影记录作为测试模型效果的数据集合,共27 482条记录。

3.2 基于用户决策机理的电影推荐实验过程

根据前文构建的推荐模型,基于用户决策机理的电影推荐实验主要包括用户观影决策机理分析、特征选取、兴趣建模和推荐结果生成几个环节。

3.2.1 用户观影决策机理分析 豆瓣电影的用户以大学生、年轻白领等高学历的年轻人为主,而且用户的观影兴趣较为稳定,因此为分析用户的电影决策机理,选取了15位华中师范大学和武汉大学的在校学生进行了访谈。总体而言,用户观影决策中考虑的主要因素较为一致,包括演员、导演、主题/类型、国家地区、评分、流行度、新颖度几个因素。其中,演员、导演、主题/类型、国家地区没有明显的作用倾向,需要根据用户的个人偏好而定;评分、流行度和新颖度则有明显的作用倾向,用户更喜欢评分较高、较为流行和新颖的影片,但其作用大小则受用户个人偏好的影响。需要说明的是,一部影片的主题/类型、演员、导演、国家地区的取值可能都不止一个,在决策过程中,用户对每个因素的兴趣度往往并非其各个取值兴趣度的简单相加,而更可能取各特征取值的兴趣度最大值。针对不同的影片,这些因素间的作用机制有所区别,常见的有以下两种:①如果特别喜欢影片的导演或主要演员,而且评分、流行度、新颖度在可接受范围内,则会选择观看;②如果主要演员或导演不熟悉、不够喜欢,则会综合考虑主题/类型、国家地区、评分、流行度、新颖度5个因素进行决策,在实际决策中,往往先依据主题/类型进行初步过滤。

3.2.2 特征选取 尽管前面所述的影响用户决策的因素之间存在一定的相关性,如导演和演员与影片的主题/类型、国家地区、流行度是显著相关的,但用户在决策过程中将它们视为不同的因素,因此在进行电影特征选择时与这些因素相关的特征都被考虑进来。具体而言,选择的特征包括(特征与影响因素间的对应关系见表1):①导演和演员,这两个因素发挥作用的方式及对用户的影响程度相近,且取值上存在一定的重合,因此将其归并到一起,统称"创作者";②主题/类型,影片的主题与类型存在一定的交叉,因此也将其放到一起,如动作、家庭、警匪等;③制片国家地区,用户决策中考虑的国家地区是综合主要演员、影片故事发生地或影片主要人物所属区域进行判断的,但这一信息难以获得,考虑实际情况,采用影片的制片国家地区进行代替;④评分,不同人群对影片的评分可能有较大

差异,本研究中采用豆瓣用户的评分;⑤观影人数,其依据是用户决策中考虑的流行度是该影片是否被广大用户观看,这一信息可以用豆瓣电影页面上的观影人数反映;⑥上映时间,该特征的作用是用于判断影片在当前或某个特定时刻的新颖度。

表 1　建模选用特征与用户决策影响因素间的映射

所选特征	决策影响因素
创作者	演员、导演
主题/类型	主题/类型
制片国家地区	国家地区
豆瓣评分	评分
观影人数	流行度
上映时间	新颖度

3.2.3　兴趣建模　为建立用户电影兴趣模型,首先需要对观影记录和影片基本数据进行预处理,在此基础上计算用户对每个特征项的兴趣度、生成用户决策函数,并对其进行语义描述。

(1)数据预处理。该部分工作主要包括以下几个方面:①影片的创作者信息提取。包括影片的导演和演员表中排名前 4 的演员(从历届奥斯卡金像奖、金鸡奖、百花奖、台湾金马奖和香港金像奖来看,最佳男女主角几乎都位于演员表的前 3 位,同时考虑到存在少量双生双旦和多主角电影,因此取排序前 4 的演员),并且对同一部影片中的导演和演员去重。②影片主题/类型信息提取。该类信息通过用户添加的标签进行提取,实现上采用了文献[24]的方法进行相关主题/类型标签的识别,并且剔除了相关影片少于 50 部的冷门主题/类型。③从抓取的“制片国家地区”字段提取影片的国家地区信息。④新颖度提取与离散化。新颖度是一个相对概念,在对用户兴趣度计算时,需要根据其观影时间与上映时间之差来衡量其新颖度。为便于分析,依据 830 682 用户 2013 年观影数据将新颖度离散化:首先,将数据以半年为区间进行分割,并统计用户观影量的分布;其次,统计结果显示,上映半年以内的影片占用户观影量的 30.7%,7－12 个月的占 7.0%,明显高于其他区间,据此,将上映半年以内、6－12 个月作为新颖度的两个区间;再次,1－3 年、3－5 年、5－10 年这几个大的时间区间内,各半年区间的数据分布较为接近,如 5－10 年区间的数据均分布在 1.45%－1.81% 之间,据此将上映 1－3 年、3－6 年、6－10 年作为新颖度的 3 个区间;最后,上映超过 10 年的影片,其被观看的比例相对较低,为简化计算,不再细分,将其作为一个区间。⑤评分提取与特征离散化。如果抓取的评分字段非空,则以其取值作为影片评分,否则采用影片平均分 6.9 分作为其评分。此外,豆瓣对电影的评分采用 10 分制,为简化计算,将其分为 5 分及以下、5.1－6 分、6.1－7 分、7.1－8 分、8 分以上 5 个区间。⑥流行度信息离散化。流行度的取值范围是一个很大的连续区间,为便于衡量,采用如下方法进行处理:按照观影频次从高到低对影片排序,并将累计观影频次占全部观影频次 40% 的影片作为第 1 档;对于剩余的影片,按照相同方法进行处理,最终将其分为 10 档。预处理后的影片信息如表 2 所示:

表 2　预处理后的影片数据(局部)

影片名	创作者	主题/类型	国家地区	新颖度	评分	流行度
大鱼	伊万・麦克格雷格,杰西卡・兰格,比利・克鲁德普,蒂姆・波顿	剧情,奇幻	美国	10 年以上	8 分以上	流行度 1 档
11 时	崔丹尼尔,李代延,郑在咏,金炫锡	科幻	韩国	半年内	6.1－7 分	流行度 6 档
扫毒	刘青云,古天乐,张家辉,袁泉,陈木胜	动作,犯罪	香港	半年内	7.1－8 分	流行度 1 档
就是闹着玩的	卢卫国,李易祥,王彤	喜剧	中国	1－3 年	6.1－7 分	流行度 5 档
毒战	古天乐,孙红雷,杜琪峰,钟汉良,黄奕	犯罪,警匪	香港	1－3 年	7.1－8 分	流行度 1 档
流感	张赫,柳海真,秀爱,金成洙	灾难	韩国	6－12 个月	7.1－8 分	流行度 3 档
巴菲的奇妙命运	伊利亚娜・狄克鲁兹,佩丽冉卡・曹帕拉,兰比尔・卡普尔	人生,喜剧,爱情	印度	1－3 年	8 分以上	流行度 4 档
赫尔克里的丰功伟绩	埃莉诺・汤姆林森,大卫・苏切特,安迪・威尔逊,鲁珀特・伊文斯	悬疑,犯罪	英国	半年内	8 分以上	流行度 8 档

(2)兴趣度计算。兴趣度是为推荐结果生成环节计算候选对象与兴趣模型匹配度服务,在计算策略设计中需要考虑后续的应用需求。在电影决策中,用户一般首先根据创作者或主题/类型进行初步决策,然后根据国家地区、评分、流行度、新颖度做出最终决策。基于此,创作者、主题/类型兴趣度的内涵应当是,给定一部包含该特征值的影片,用户愿意观看的概率,因此,这两个特征的兴趣度取值范围应是[0,1];剩余 4 个特征的兴趣度内涵应当是,给定一部包含该特征值的影片,相较于该特征的其他取值,用户在多大程度上

更愿意看该影片，因此其取值应围绕 1 进行波动，理论上范围为[0，+∞）。以此出发，可以采用公式（1）计算创作者、主题/类型兴趣度，公式（2）计算其他 4 个特征的兴趣度。

$$W(u_i, t_j) = \frac{F(u_i, t_j) - 1}{F(t_j)} \qquad 公式(1)$$

$$W(u_i, t_j) = \frac{F(all) * (F(u_i, t_j) - 1)}{F(u_i) * F(t_j)} \qquad 公式(2)$$

其中，$W(u_i,t_j)$指用户 u_i 对特征值 t_j 的兴趣度；$F(u_i,t_j)$指用户观看过的包含特征值 t_j 的影片数量，为避免随机因素的影响，要求 $F(u_i,t_j) \geqslant 3$；$F(t_j)$ 指包含特征值 t_j 的影片总数；F(all)指影片总数；$F(u_i)$ 指用户 u_i 看过的影片总数。

（3）用户决策函数拟合。依据用户决策中各因素间的关联关系，可以将用户决策模式进一步抽象为：①依据影片的创作者或主题/类型和国家地区进行初步决策，如果感兴趣，则进入下一步；②依据影片的新颖度、评分和流行度进一步决策。故而，可以将用户决策函数的形式表示为公式（3）所示的形式（为平滑影响，对国家地区、新颖度、评分和流行度的影响进行了对数处理）。

$$W(u_i,m_j) = \begin{cases} max(maxw(u_i,cre_j),maxw(u_i,typ_j) * (1+maxlogw(u_i,cou_j))) * (1+logw(u_i,rat_j)) * (1+logw(u_i,popt_j)) * (1+logw(u_i,nov_j)) \\ (if\ logw(u_i,rat_j) > -1, logw(u_i,pop_j) > -1, logw(u_i,nov_j) > -1) \\ 0 \quad (if\ logw(u_i,rat_j) \leqslant -1,\ or\ logw(u_ipop_j) \leqslant -1, or\ logw(u_i,nov_j) \leqslant -1) \end{cases}$$

公式（3）

其中，$W(u_i,m_j)$表示影片 m_j 与用户 u_i 的兴趣模型的匹配度，$w(u_i,cre_j)$、$w(u_i,typ_j)$、$w(u_i,cou_j)$、$w(u_i,rat_j)$、$w(u_i,pop_j)$、$w(u_i,nov_j)$表示用户 u_i 对影片的创作者 cre_j、主题/类型 typ_j、国家地区 cou_j、评分 rat_j、流行度 pop_j、新颖度 nov_j 的兴趣度。

（4）语义描述。由于用户观影决策中考虑的影响因素较为一致，因此，每个用户的语义描述框架都是由{创作者，主题/类型，国家地区，新颖度，评分，流行度}6 个要素组成的高维向量；其中每一个维度都由一系列的特征值及兴趣度二元组，以创作者维度为例，其可以表示为［创作者：（刘德华，0.34），（李冰冰，0.13）……］，示例如表 3 所示：

表 3 用户语义兴趣模型（局部）

用户 ID	创作者	主题/类型	国家地区	新颖度	评分	流行度
1065097	安贝・瓦莱塔，0.40	黑色幽默，0.039	香港，1.172	半年内，1.022	8 分以上，1.176	流行度 1 档，1.830
11542932	姜孝镇，0.33	警匪，0.076	韩国，1.202	6－12 个月，0.768	7.1－8 分，1.124	流行度 2 档，1.642
12489410	王川，0.25	人性，0.028	香港，1.218	1－3 年，0.819	7.1－8 分，1.021	流行度 1 档，1.939
1259677	中岛哲也，0.25	心理，0.017	泰国，1.572	半年内，1.170	8 分以上，1.228	流行度 2 档，1.662
1297671	李治廷，0.13	人生，0.028	香港，1.152	6－12 个月，0.748	7.1－8 分，1.176	流行度 3 档，1.322

3.2.4 推荐结果生成　在推荐结果生成环节，首先对候选推荐影片进行了语义表示，在创作者、主题/类型、国家地区 3 个特征权重上，采用了二元赋权的方法，即如果包含了某个特征，则取值为 1，否则为 0；对于新颖度、评分和流行度 3 个特征，则根据其实际情况映射到对应区间；在此基础上，将其表示为向量形式。其次，按照用户的个人兴趣模型及决策函数，计算了每部候选推荐影片的推荐度。最后，将候选推荐影片按照推荐度进行降序排列，并选取权值最高的 N 个作为最终的推荐结果。在实验中，N 分别取 5、10、20、50、100 和 200。

3.3 基于向量空间模型的电影推荐实验过程

在研究中，基于向量空间模型的电影推荐是对照实验，用于作为基准线评判前一个实验的效果。在实验设计上，综合基于内容的电影个性化推荐相关研究成果[25-27]，选取导演、演员、主题/类型、国家地区作为特征，并在推荐结果生成中利用流行度对结果进行调权。

在兴趣建模中，假设 $w(u_i,t_j)$表示用户 u_i 对特征值 t_j 的兴趣度，$F(u_i,t_j)$表示特征值 t_j 在用户 u_i 看过影片中出现的频次，$F(u_i)$指用户 u_i 看过的影片数量，则可以采用公式（4）计算每一个特征值的兴趣度。

$$w(u_i, t_j) = \frac{F(u_i, t_j)}{F(u_i)} \qquad 公式(4)$$

流行度调权策略设计上采用了与前一个实验相似的思路，按照同样的区间划分方式将电影流行度分成 10 个区间，并假设 $w_{pop\text{-}j}$ 表示流行度区间 j 的权重，F(all)指影片总数；F(pop-j)指流行度区间为 j 的影片累计观影频次占总观影频次的比例；N(pop-j)指流行度区间为 j 的影片数量，则流行度区间 j 的权重可以通

过公式(5)进行计算。

$$w_{pop-j} = 1 + \log \frac{F(all) * F(pop-j)}{N(pop-j)} \quad 公式(5)$$

在确立兴趣度计算公式和流行度调权策略的基础上，可以通过用户兴趣向量 $u_i((t_1, w(u_i, t_1)), ((t_2, w(u_i, t_2)) \cdots ((t_j, w(u_i, t_j)) \cdots)$、影片特征向量 $m_j((f_1, 1), ((f_2, 1) \cdots ((f_k, 1) \cdots)$ 及对应的流行度权值 w_{pop-m_j} 的积来计算影片与用户兴趣模型的最终匹配度，如公式(6)所示。

$$W(u_i,\ m_j) = u_i * m_j * w_{pop-m_j} \quad 公式(6)$$

3.4 实验结果及分析

在推荐实现中，以全部75 694部影片作为候选集，但对于每一位用户，则剔除其于2014年5月3日以前已经看过的影片。为便于衡量实验效果，选取较为典型的P@N作为评价指标[28]，并采用卡方检验进行显著性检验，结果如表4所示；

表4 实验结果及其显著性检验

推荐方法	P@5	P@10	P@20	P@50	P@100	P@200
用户决策机理	0.53%***	0.54%***	0.47%***	0.41%***	0.34%***	0.28%***
向量空间模型	0.18%	0.21%	0.24%	0.23%	0.23%	0.22%

注：***表示 $p < 0.001$

从表1可以看出，基于用户决策机理的个性化推荐模型效果显著好于向量空间模型的效果($p < 0.001$)，而且从直观上看，推荐效果提升也非常显著，以P@10和P@20为例，基于用户决策机理的推荐模型的准确率分别为0.54%和0.47%，与之相对的，向量空间模型分别为0.21%和0.24%。实验模型效果更优的根本原因是模型所采用的思路更好地拟合了用户决策机理，具体而言主要包括3个方面：①从用户决策机理出发进行特征选择，全面、系统地涵盖了用户决策时的主要考虑因素；②针对每个特征维度的特点进行了权值计算方法设计，使其更符合客观实际，例如在计算用户对创作者兴趣度大小时考虑了其作品数量因素；③基于多维度融合的综合权值计算策略更符合用户决策中各个维度的关联关系。

此外，原型实验虽然验证了基于用户决策机理的个性化推荐这一思路的有效性，但实验设计仍存在一些有待于优化的问题：①在实践中，用户的决策模式更加复杂，且受决策风格的影响，不同用户间存在显著差异，这些问题在实验设计中考虑不够细致；②原型实验中的部分参数依据经验进行设置，缺乏更细致的调优环节。

4 结论

为改善基于内容的个性化推荐效果，应立足于用户决策机理进行特征选择、兴趣建模和推荐结果生成，以此出发，笔者构建了基于用户决策机理的个性化推荐模型，并以电影数据为例对模型效果进行了检验。结果显示，相对于向量空间模型，该策略能够大幅提升个性化推荐的效果，从而验证了基于用户决策机理的个性化推荐这一思路的有效性。结合电影推荐原型实验，为推动研究的深化，今后拟重点关注如下问题：①用户实际决策中因素之间的关系可能比较复杂，比如同一个特征既存在正向作用的特征值也存在负向作用的特征值，多个特征同时存在竞合关系等，针对这些问题，需要研究如何在模型构建中进行决策因素复杂关联关系的分析与应用；②不同决策风格用户关注的影响因素及利用这些因素的方法可能差异显著，为获得更好的推荐效果，就需要研究如何进行用户决策风格的自动识别及推荐模型的自适应调整；③扩展应用领域，将其应用于图书、学术论文等学术信息资源的推荐中，以进一步验证模型的效果。

参考文献：

[1] LU J, WU D, MAO M, et al. Recommender system application developments: a survey[J]. Decision support systems, 2015, 74(C):12－32.
[2] 阮光册，夏磊. 互联网推荐系统研究综述[J]. 情报学报，2015，34(9):999－1008.
[3] KUNAVER M, Požrl T. Diversity in recommender systems - a survey[J]. Knowledge-based systems, 2017, 123(9):154－162.
[4] HARIRI N, MOBASHER B, BURKE R. Context adaptation in interactive recommender systems[C]//ACM conference on recommender systems. New York:ACM, 2014:41－48.
[5] 王东山. 消费者购买决策理论评述与展望[J]. 商业经济研究，2017(21):43－46.
[6] 李宗伟，张艳辉，栾东庆. 哪些因素影响消费者的在线购买决策[J]. 管理评论，2017，29(8):136－146.
[7] LAY-YEE K L, HAN K S, FAH B C Y. Factors affecting smartphone purchase decision among malaysian generation Y[J]. International journal of Asian social science, 2013, 3(12):2426－2440.
[8] SUJATA J, YATIN J, ABHIJIT C, et al. Factors affecting smartphone purchase among Indian youth: a descriptive analysis[J]. Indian journal of science & technology, 2016, 9(15):1－10.
[9] JOO S, CHOI N. Factors affecting undergraduates' selection of online library resources in academic tasks[J]. Library Hi Tech, 2015, 33(2):114－117.
[10] 查先进，李力，严亚兰，等. 数字图书馆环境下信息有用性和

信息获取影响因素研究[J]. 情报学报, 2017, 36(7):669 - 681.
[11] 吴江, 周露莎. 网络健康信息服务用户购买决策的影响因素研究[J]. 情报学报, 2017, 36(10):1058 - 1065.
[12] BOGDANOV D, HARO M, FUHRMANN F, et al. Semantic audio content-based music recommendation and visualization: based on user preference examples[J]. Information processing & management, 2013, 49(1):13 - 33.
[13] 谭学清, 何珊. 音乐个性化推荐系统研究综述[J]. 现代图书情报技术, 2014, 30(9):22 - 32.
[14] 洪文, 聂延平, 青巧. 馆藏资源自动推荐模型结构与处理流程优化分析[J]. 情报理论与实践, 2016, 39(5):130 - 133.
[15] SOARES M, VIANA P. Tuning metadata for better movie content-based recommendation systems[J]. Multimedia tools and applications, 2015, 74(17):7015 - 7036.
[16] JUN S, KIM D, JEON M, et al. Social mix: automatic music recommendation and mixing scheme based on social network analysis [J]. Journal of supercomputing, 2015, 71(6):1933 - 1954.
[17] CHEN C, CHEN C, CHEN H, et al. Towards context-aware social recommendation via individual trust[J]. Knowledge-based systems, 2017, 127(C):58 - 66.
[18] ÅMAN P, LIIKKANEN L A. Interacting with context factors in music recommendation and discovery[J]. International journal of human-computer interaction, 2017,33(3): 165 - 179.
[19] 李江, 李东, 冯培桦, 等. 基于专长吻合度、学术影响力与社会关联值的专家推荐模型研究[J]. 情报学报, 2017, 36(4): 338 - 345.
[20] 杨程, 范强, 王涛, 等. 基于多维特征的开源项目个性化推荐方法[J]. 软件学报, 2017, 28(6):1357 - 1372.
[21] 周建中, 徐芳. 国立科研机构同行评议方法的模式比较研究[J]. 科学学研究, 2013, 31(11):1642 - 1648.
[22] 庄锦英. 决策心理学[M]. 上海:上海教育出版社,2006:12 - 16.
[23] HAN J, KAMBER M, PEI J. 数据挖掘概念与技术[M]. 3 版. 范明,孟小峰, 译. 北京:机械工业出版社, 2012:55 - 79.
[24] 林鑫, 石宇, 周知. 基于相对频次的标签相关性判断优化研究[J]. 图书情报工作, 2016,60(17):130 - 135.
[25] 吕学强, 王腾, 李雪伟, 等. 基于内容和兴趣漂移模型的电影推荐算法研究[J/OL]. [2017 - 11 - 21]. http://www.aroc-mag.com/article/02-2018-03-049.html.
[26] ÇATALTEPE Z, ULUYAGMUR M, TAYFUR E. Feature selection for movie recommendation[J]. Turkish journal of electrical engineering & computer sciences, 2016, 24(3):833 - 848.
[27] SHI C, LIU J, ZHUANG F, et al. Integrating heterogeneous information via flexible regularization framework for recommendation [J]. Knowledge & information systems, 2015, 49(3): 835 - 859.
[28] 胡潜, 林鑫. 社会化标注系统中基于标签和项目的兴趣建模比较研究[J]. 情报学报, 2015, 34(12):1296 - 1303.

作者贡献说明:
林鑫:负责数据处理,论文撰写;
桑运鑫:参与数据处理,负责论文修改;
龙存钰:负责论文修改。

Personalized Recommendation Based on User Decision-making Mechanism

Lin Xin[1,2] Sang Yunxin[3] Long Cunyu[2]
[1] Institute of Scientific and Technical Information of China, Beijing 100038
[2] School of Information Management of Central Normal University, Wuhan 430079
[3] School of Information Management of Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] The purpose of this paper is to propose an optimization strategy of features choosing and weight computing for content-based personalized recommendation. [**Method/process**] This paper proposes a personalized recommendation model based on user's decision-making mechanism, which takes user decision mechanism as background knowledge in features selection, user interest profile construction and semantic representation, and user decision function construction. To test this model, this paper conducts an experiment taking 4 748 users as sample, vector space model as reference model, and P@N as evaluation index. [**Result/conclusion**] The results show that, in the cases of N equals 5, 10, 20, 50, 100, 200, the personalized recommendation model based on user decision-making mechanism is significantly better than the vector space model, and the effectiveness of the model is verified.

Keywords: decision-making mechanism content-based recommendation personalized recommendation